Projet ORACLE – Task 4.2 - Note technique

Modèles d'impact climatique pour la productivité et la probabilité de présence d'espèces ligneuses principales de la forêt française

20 janvier 2014 – corrigé janvier 2016 (v4)

Pierre Mérian,

INRA, Centre de Nancy-Lorraine, UMR 1092 LERFoB
IGN, Laboratoire de l'Inventaire Forestier

Jean-Daniel Bontemps

AgroParisTech, Centre de Nancy, UMR 1092 LERFoB

Correspondance :

jdbontemps.agroparistech@gmail.com

1 Projet

La tâche assignée est de développer les outils qui permettent de rendre le modèle de dynamique de la ressource forestière (LSFDM pour Large Scale Forest Dynamics Model, **Wernsdörfer et al. 2012**) dépendant des facteurs de l'environnement, et en particulier des facteurs du climat, afin de réaliser des projections à l'horizon 2100. Dans le projet, ce modèle est couplé au FFSM (French Forest Sector Model ; **Caurla et al. 2010**, document de travail), qui est un modèle d'équilibre partiel du marché forestier, dans lequel les décisions de prélèvement (ou de reboisement) résultent de décisions économiques. La dépendance à l'environnement concerne la croissance à cortège ligneux donné (feuillus ou résineux), et les éventuelles transitions de cortèges ligneux. Pour cela, l'idée est de s'appuyer sur les modèles environnementaux de productivité (aspect croissance), et de distribution (aspect transition de cortèges), développés, à l'échelle de l'espèce ligneuse, dans l'équipe « écologie forestière ».

Le projet est une expertise intégrative, et vise à des projections conditionnellement à l'état de l'art. Les solutions proposées dans le projet sont donc des solutions pragmatiques (logique d'ingénierie), et réalistes autant que possible, mais peuvent appeler des questionnements scientifiques, dont la résolution est laissée de côté dans le cadre du projet.

Il s'agira dans un premier temps de comprendre le déterminisme environnemental de la distribution et de la productivité des principales essences forestières françaises, sur l'ensemble de la France. Pour cela, des modèles de distribution et de productivité seront ajustés par essence sur le territoire national. Ces modèles, dont une partie des prédicteurs seront climatiques, seront ensuite projetés sous climat futur à l'horizon 2100.

Il s'agira ensuite de produire par région administrative et par groupe d'essences (feuillus/résineux) des modificateurs des paramètres (taux de passage entre classes de diamètre) du modèle LSFDM afin que ce dernier soit sensible au climat. Ce modèle est détaillé dans la publication Wernsdörfer et al (2012) et dans la note « Programme de travail Oracle – équipe EF » du 13 juin 2012 (**Bontemps 2012**).

2 Données

2.1 Données forestières

Les données de présence/absence (modélisation de la distribution) et d'accroissement en surface terrière (modélisation de la productivité) sont issues des **campagnes IFN de 2005 à 2009**. Il s'agit de données « nouvelle méthode » téléchargeables depuis le site de l'Inventaire Forestier.

Dans l'équipe Écologie Forestière (EF), ces données sont téléchargées et mises en forme en BDD par Ingrid Seynave. Les données IFN utilisées dans le projet ont donc été extraites de cette BDD, pour un total de **33471 points**. A chaque point sont associées diverses données, dont notamment :

- Des données relatives à la placette (localisation, topographie, sol, etc.) ;
- Des données relatives au peuplement (essence principale, taux de couvert, taille du massif, etc.) nécessaires aux sélections de placettes (cf. 3.2) ;
- Un relevé floristique, à partir duquel les présences/absences sont obtenues pour ajuster les modèles de distribution ;
- Un accroissement en surface terrière (m²/ha/an) calculé par Ingrid Seynave à partir des données brutes d'accroissement de l'IFN (IR5).
- De nouvelles variables nécessaires à la sélection de placettes et/ou à la modélisation de la productivité ont été calculées par Ingrid Seynave, notamment la hauteur dominante (H0), l'*indexe relatif de densité* (RDI) et la ventilation de la surface terrière entre essences (cf. 3.2).

Notons ici que, lors de la campagne de l'année N, la présence/absence d'une espèce est observée directement sur la placette (année N), alors que les données d'accroissement correspondent à la croissance des 5 années précédentes (N-1 à N-5). **La présence/absence est définie à partir de la réunion des observations dendrométriques (diamètre de recensement de 7.5 cm) et d'observations du relevé floristique (stades juvéniles depuis le semis, et non recensables). A ce titre, la présence/absence est la conséquence de conditions environnementales sur une période indéfinie (de quelques années pour les semis à quelques décennies pour les arbres mâtures**), alors que l'accroissement sur 5 ans dépend des conditions environnementales des cinq années correspondantes. Un IR5 obtenu lors de la campagne de l'année N dépend des conditions environnementales des années N-1 à N-5.

2.2 Données climatiques

2.2.1 Climat passé

Les données climatiques passées SAFRAN (1958-2010) ont été téléchargées sur le portail HYMEX. Les données SAFRAN sont des données horaires couvrant la France à une **résolution de 8 km sur une projection Lambert-II étendue** (**Pagé 2008**). Elles sont produites par Météo-France (Centre National de Recherches Météorologiques, CNRM). Une description du système SAFRAN appliqué à la France

entière est décrite dans **Le Moigne (2002)**. Dans le cadre du projet ORACLE, les données horaires ont été converties pour chaque année en données mensuelles et saisonnières.

2.2.2 Climat futur

Les données climatiques futures CERFACS (2000-2100) ont été obtenues auprès du CERFACS qui est partenaire de l'ANR Oracle. Au CERFACS (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique), les données SAFRAN sont utilisées dans le cadre de désagrégation de simulations climatiques afin de produire des données climatiques haute-résolution couvrant la France sur la grille SAFRAN. A la rédaction de cette note, 15 combinaisons « GCM x Scénario x Membre » sont disponibles. Ces 15 combinaisons sont réparties comme suit et visent à évaluer, suivant un plan d'expérience partiel emboîté, les différences sources d'incertitude relatives aux simulations des GCM (épistémique, réflexive, et stochastique ; **Pagé (2011)** = scénarios Scratch) :

GCM / scén	a1b	a2	b2
arpege	1	1	1
CNCM33	1	/	/
DMIEH5C	3	/	/
EGMAM2	1	/	/
HADGEM2	1	/	/
IPCM4	3	/	/
MPEH5C	3	/	/

Tableau 1. *Nombre de membres disponibles par combinaison GCM x Scénario*

2.2.3 Variables climatiques disponibles

Toutes les variables disponibles ont été extraites. Une partie d'entre elles a été recalculée afin d'obtenir des unités cohérentes avec l'exploitation des données telle qu'envisagée au LERFoB (par exemple, les précipitions en $kg.m^{-2}.s^{-1}$ sont converties en mm de précipitation). Les données brutes SAFRAN et CERFACS contiennent les variables suivantes :

Nom variable	Définition	Unité CERFACS	Unité après conversion
PRCP	Précipitations liquides	$Kg.m^{-2}.sec^{-1}$	mm
SNOW	Précipitations solides	$Kg.m^{-2}.sec^{-1}$	mm
Q	Humidité spécifique	$Kg.kg^{-1}$	$Kg.kg^{-1}$
Vu	Vitesse du vent à l'horizontal à 2 m	$m.s^{-1}$	$m.s^{-1}$
T	Température à 2 m	°C	°C
GLO	rayonnement visible incident à la surface	$W.m^{-2}$	$W.m^{-2}$
RAT	rayonnement infrarouge incident	$W.m^{-2}$	$W.m^{-2}$

Tableau 2. *Variables climatiques disponibles dans les analyses SAFRAN/CERFACS.*

A partir de ces variables, d'autres variables ont été calculées pour obtenir la liste finale suivante :

Variable	Description	Unité
GLO	Incoming Solar Radiation (spectre visible)	$W.m^{-2}$
HR	Humidité relative moyenne mensuelle	%
PRCP	Précipitation liquide mensuelle	mm
Psat	Pression de vapeur d'eau saturante mensuelle	hPa
Ptot	Précipitation totale (liquide + solide)	mm
Q	Humidité spécifique moyenne mensuelle	$kg.kg^{-1}$ ↔ s. u.
RAT	Radiation infrarouge (spectre infrarouge)	$W.m^{-2}$
SNOW	Précipitation solide mensuelle	mm
T	Température moyenne mensuelle	°C
Tmax	Température maximale mensuelle	°C
Tmin	Température minimale mensuelle	°C
VPD	Déficit de pression de vapeur	hPa
Vu	Vent à 2 m du sol	$m.s^{-1}$

Tableau 3. *Liste finale des variables climatiques disponibles à partir des analyses SAFRAN/CERFACS. Le détail du calcul des variables HR, Psat et VPD est fourni en Appendice 1.*

2.3 Données édaphiques

Initialement, les données édaphiques étaient de plusieurs origines :

- Relevées sur le terrain : la profondeur du sol (variant de 1 à 9 dizaines de cm) ;
- Bio-indiquées (estimées à partir du relevé floristique) : pH, CN (**Gégout et al. 2003**). *Note : les engorgements temporaires et permanents ont été retirés du projet car ils présentaient un biais d'absorption d'effets climatiques et de fertilité du sol à échelle spatiale pluri-kilométrique ; ces variables permettaient certes d'augmenter le pouvoir descriptif des modèles mais privaient l'entrée de variables climatiques et/ou édaphiques dans les modèles nécessaires aux interprétations écologiques fines.*
- Extraites de couches spatialisées du LERFoB : RUM (réserve utile maximale) obtenu à partir de la couche *rumkg_500* (voir Christian Piedallu ou Vincent Perez pour plus de détails).

Contrairement aux données climatiques SAFRAN et CERFACS qui sont spatialisées à une maille de 8km, les données édaphiques sont ponctuelles (une donnée par placette IFN). **Cette différence de résolution spatiale des données peut engendrer un biais de modélisation, car la variabilité intra-maille de la productivité ou de la distribution des espèces dans la maille SAFRAN de 8 km ne peut être expliquée que par les variables édaphiques. Les données édaphiques fournies par Ingrid Seynave ont ainsi été abandonnées au profit de couches spatialisées produites par le LERFoB (cn_kg_2007 pour le CN, ph_kg_2008 pour le pH et rumkg_500 pour la RUM, Piedallu et al. 2013)**. Toutes les couches ont été produites par Christian Piedallu à partir des données IFN ancienne méthode (environ 140 000 points). La méthode suivie est décrite par Christian Piedallu dans un guide technique publié en 2008 (AgroParisTech-ENGREF (UMR LERFoB), IFN, 2008). **Les couches SIG étaient**

à une résolution de 1 km (sauf pour la RUM qui est à une résolution de 500 m) et ont été dégradées (moyenne) à la maille SAFRAN de 8km.

La profondeur du sol n'a pas été krigée par le LERFoB. **La couche SIG a été produite en utilisant les données IGN nouvelle méthode du projet Oracle (33 471 points environ) en respectant la méthode du guide technique**.

Au final, les 4 variables édaphiques sont spatialisées à l'échelle de la France et sur la maille SAFRAN projetée en Lambert-II étendu. Pour chaque placette IFN, les données édaphiques utilisées en modélisation ont été extraites de ces couches grâce aux coordonnées géographiques.

Nom variable	Définition	Unité
pH	Acidité du sol	unités pH
CN	Rapport carbone/azote	kg C / kg N
RUM	Réserve utile maximale	Mm
Prof	Profondeur du sol	**Dm**

Tableau 4. *Variables édaphiques utilisées dans le projet.*

3 Quelle forêt modélisée ?

3.1 Distribution

Les modèles de distribution sont ajustés sur les données de présence/absence des relevés floristiques de chaque placette. Nous modélisons donc ici la niche réalisée de l'essence.

3.2 Productivité

Le cadre de modélisation de la productivité est imposé par la variable Y que l'on modélise : l'accroissement en surface terrière. La productivité d'un peuplement, appréhendée par les IR5, est très sensible au stade de développement du peuplement et à la compétition : plus le peuplement est en stade avancé (mâture, sénescent), moins il sera productif ; plus le stock sur pied sera élevé (couvert fermé), plus la productivité sera élevée. Ces deux effets doivent être pris en compte dans les modèles, et sont respectivement estimés par la hauteur dominante (H0) et l'*indice de densité relative* (RDI). Les peuplements purs, réguliers fermés, avec des arbres biens conformés, localisés dans des grands massifs forestiers (> 4 ha), hors lisière sont ainsi des communautés de référence pour lesquels nous disposons d'outils pour extraire le plus proprement possible le signal environnemental de la croissance radiale. Le calcul du RDI s'est appuyé sur les équations publiées par **Charru et al. (2012)**. Le tableau ci-dessous liste les critères de sélection des placettes conservées pour l'ajustement des modèles de productivité :

Variable IFN	Définition	Critères de sélection	Peuplements conservés SSI	Fractions
sfo	Structure Forestière	1 ou -999	Futaie Régulière ou non renseigné	0 à 4
propG_taillis	Proportion des brins de taillis en surface terrière	<25%	Moins de 25% de surface terrière en brins de taillis	0 à 4
ETR_H	Ecart-type relatif des hauteurs totales individuelles	<40%	L'écart-type relatif des hauteurs doit être <40%	0 à 4
tm2	taille de massif	3	Couverture du sol "forêt" et surface >= 4ha	0 à 4
SomTCA	Sommes des taux de couverts absolus	>= 50 ou -999	Taux de couvert absolu >=50 % ou non renseigné	1 à 4
NbFORME	Nombre de tiges têtard ou à fort houppier	0	Aucune arbre têtard et aucun arbre à fort houppier	0 à 4
plisi	présence de lisière	0 ou -999	Pas de lisière ou non renseigné	2 à 4
peupnr	peuplement non recensable	0 ou -999	Peuplement recensable ou non renseigné	4
Csa	couverture du sol	1	Couvert boisé fermé	0 à 4
dc	type de coupe	0, 8, 9 ou -999	Pas de coupe ou non renseigné	2 à 4
propG_ess	proportion de la surface terrière de l'essence prépondérante	> 80%	> 80% de la surface terrière pour l'essence prépondérante	0 à 4

Tableau 5. *Critères de sélection des placettes IFN pour la définition du jeu de données « productivité ».*

Cette sélection réduit fortement le nombre de placettes IFN par essence ainsi que les gradients écologiques couverts par ces placettes. 8 essences avaient été initialement pressenties pour l'analyse. Cependant, dans le cas du pin d'Alep, les gradients climatiques présentaient de trop fortes corrélations pour être identifiés sans confusion, et l'espèce a été écartée.

Au final, 7 essences ont été conservées :

- *Abies alba* (sapin pectiné) : **Aa → 392 placettes**
- *Fagus sylvativa* (hêtre commun) : **Fs → 496 placettes**
- *Picea abies* (épicéa commun) : **Pa → 571 placettes**
- *Pinus sylvestris* (pin sylvestre) : **Ps → 751 placettes**
- *Quercus petraea* (chêne sessile) : **Qpt → 608 placettes**
- *Quercus pubescens* (chêne pubescent) : **Qpb → 155 placettes**
- *Quercus robur* (chêne pédonculé) : **Qr → 458 placettes**

4 Procédures de sélection pré-modélisation

Elles visent à obtenir un jeu de données « propre » sur lequel pourront être ajustés les modèles. Trois points ont été traités simultanément et conjointement à la phase de modélisation (bien que présentés successivement ci-après) :

- Limiter les colinéarités entre variables ;
- Obtenir un pool générique et parcimonieux de variables environnementales ;
- Éliminer les *outliers*.

4.1 Sélection des variables environnementales

A ce stade du travail, nous disposons de 4 variables édaphiques (pH, CN, RUM et profondeur du sol) et de 85 variables climatiques (températures moyennes, minimales et maximales, précipitations, déficit de pression de vapeur) à des résolutions temporelles différentes (année, saison, mois). Une procédure de sélection des variables, notamment climatiques, est nécessaire.

Notons ici que les variables climatiques ont toujours été considérées selon l'année biologique et non calendaire. L'année biologique N est fixée de septembre N-1 à août de l'année N. Notons également que **ces variables climatiques diffèrent entre distribution et productivité** :

- **Distribution** : valeurs moyennes sur la période trentenaire **1971-2000** ;
- **Productivité** : comme la productivité est estimée à partir des IR5, il s'git des valeurs moyennes sur la période de 5 ans précédent l'année de la campagne. **Ainsi, les campagnes 2005 à 2009 sont respectivement confrontées au climat moyen sur les périodes allant de 2000-2004 à 2004-2008.**

4.1.1 La colinéarité comme critère de sélection

Trois cas de figure sur les corrélations de Pearson sont à distinguer dans la sélection des variables :

- 0.7 (R^2 = 0.49), en dessous duquel les variables sont considérées comme peu corrélées et non sujettes à sélection ;
- 0.8 (R^2 = 0.64), en dessus duquel les variables sont considérées comme corrélées et soumises à sélection.
- [0.7 ; 0.8], plage soumise à des tests et des réflexions (traitement au cas par cas).

4.1.2 Variables climatiques

Quelles variables climatiques ?

La première sélection porte sur la définition et conservation de groupes de variables. **Trois groupes de variables climatiques ont été définis : les températures moyennes (T), les précipitations (P) et les déficits de pression de vapeur (VPD)**. Les températures minimales et maximales mensuelles ont été exclues car très corrélées avec la température moyenne (R > 0.95).

Quelle résolution temporelle ?

L'analyse des colinéarités a montré que les variables mensuelles sont corrélées aux variables saisonnières (R > 0.975 pour T et > 0.95 pour P et VPD). **L'échelle mensuelle a été ainsi abandonnée car redondante avec l'échelle saisonnière.**

La colinéarité entre le climat entre deux saisons non consécutives (e. g. hiver-été, printemps-automne) peut être faible notamment pour P et VPD, avec des R compris entre 0.4 et 0.9 (0.6 à 0.95 pour T). **Il a donc été conclu que le pas de la saison était le plus pertinent**, l'année ne permettant pas de refléter les différences entre saisons non consécutives**. A cette étape, nous avons donc 12 variables climatiques : [Tmoy, P, VPD] x [4 saisons].**

Quelles variables physiologiquement pertinentes ?

Une dernière sélection a été faite sur d'une part le sens physiologique des variables climatiques et sur leurs colinéarités. **L'analyse des colinéarités inter-saisons conduit à ne retenir que deux saisons non-consécutives parmi quatre. Le couple hiver-été a été préféré à automne-printemps** car il reflète mieux les extrêmes climatiques de l'année biologique.

Le VPD hivernal a été retiré de l'analyse car 1) ce déficit est proche de zéro, 2) les essences ne poussent pas en cette période et 3) le VPD ne contribue pas à remplir des réserves hydriques du sol contrairement aux précipitations.

Pool climatique

Au final, le pool contient 5 variables : Twin, Tsum, Pwin, Psum et VPDsum (win = winter, sum = summer).

4.1.3 Variables édaphiques

La corrélation entre les 4 variables édaphiques souvent sous le seuil de R de 0.5, à l'exception du couple pH-CN. **Le pH est alors privilégié sur le rapport C:N** car il s'agit d'une variable (1) facilement interprétable, (2) plus fréquemment utilisée (et connue des autres acteurs du projet ORACLE) et, (3) mieux mesurée en laboratoire (le CN est un rapport de deux mesures alors que le pH résulte d'une seule mesure).

4.2 Pool générique de variables et cas particuliers

Le pool initial est de 9 variables : **Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof.** Dans certains cas, il est nécessaire soit de retirer une variable, soit d'être vigilant lors de la modélisation :

- **Modèles de distribution (toutes les essences)** : vigilant sur Twim-Tsum (R = 0.80) et Tsum-VPDsum (0.69) ;
- **Modèle de prod. Aa** : vigilant sur Twim-Tsum (0.68) et Tsum-VPDsum (0.74) ;
- **Modèle de prod. Fs** : **retrait du CN** ;
- **Modèle de prod. Pa** : vigilant sur Twim-Tsum (0.78) et Tsum-VPDsum (0.71) ;
- **Modèle de prod. Ps** : **retrait du CN** ; vigilant sur Twim-Tsum (0.79) et Tsum-VPDsum (0.65) ;

- **Modèle de prod. Qpt** : **retrait du CN** ; vigilant sur Tsum-VPDsum (0.82) ;
- **Modèle de prod. Qpb** : vigilant Twim-Tsum (0.78) et Tsum-VPDsum (0.64) ;
- **Modèle de prod. Qr** : vigilant sur Tsum-VPDsum (0.66).

4.3 Retrait des *outliers*

La détection et le retrait des *outliers* est une étape cruciale, notamment dans le cadre d'ajustement de modèles additifs généralisés (GAM – cf. 5). La forme générale des effets est sensible aux points extrêmes, qu'ils soient sur la variable Y ou sur les variables X. Une procédure d'élimination des extrêmes a donc été réalisée à la fois sur les prédicteurs et sur les variables prédites.

4.3.1 Méthode

La cross-validation des ajustements GAM (cf. 5) conduit à avoir des points de validation en dehors du jeu de calibration. Les extrapolations des GAM étant très hasardeuses, il faut supprimer les points les plus extrêmes des gradients écologiques afin de limiter les extrapolations lors de la validation croisée. De façon pragmatique, on a procédé ici à la suppression des Z % des valeurs extrêmes. **Attention : un taux de retrait de Z % signifie que l'on retire Z/2 % des valeurs extrêmes à chaque extrémité du gradient.**

La présence d'un *outlier* dans le jeu de données conduit à une baisse forte des statistiques de validation (R^2 de validation dans notre cas) du GAM car la valeur prédite de l'*outlier* est aberrante. Cette variation du R^2 (souvent spectaculaire, passant de valeurs comprises en 0.3 et 0.6 à des valeurs proches de 0) a été utilisée comme proxy de détection des *outliers*. Comme les variables retenues dans les modèles finaux ne pouvaient pas être connues au moment de la détection des *outliers*, la détection a été réalisée variable par variable et par essence.

4.3.2 Distribution

- Variable Y : 1/0 (présence/absence). Il n'existe donc pas d'*outliers*.
- Variables X : elles sont obtenues à partir de 33471 points, ce qui conduit à avoir de nombreux points aux extrémités des gradients écologiques. Quelques *outliers* correspondant à des valeurs de VPD mal estimées en montagne ont cependant été retirés (limite d'application des formules de calcul du VPD – cf. Appendice 1).

33445 placettes ont été conservées.

4.3.3 Productivité

- Variable Y : Y varie entre [0 ; +∞[. Il n'existe donc pas d'*outliers* à la borne inférieure. En revanche, **les 1 % valeurs les plus fortes (borne supérieure) ont été retirées**.
- Variables X : l'analyse a révélé qu'un *outlier* sur une variable X l'est souvent sur les autres variables. En conséquence, les Z % des valeurs extrêmes à retirer sont très proches entre les différentes variables. **Une seule valeur de %Z a ainsi été retenue par essence**, cette valeur étant appliquée à l'ensemble des X (donc retrait pas nécessairement cumulatif):

Essence	Aa	Fs	Pa	Ps	Qpt	Qpb	Qr
Taux de retrait	1%	1.5%	1%	1.5%	1%	1%	2%

Tableau 6. *Taux de retrait de placettes appliqués à l'issue de la procédure de détection d'outliers.*
Note 1 : ces taux de retrait sont élevés car les GAM sont très sensibles aux outliers. *Des taux élevés garantissent des statistiques de validation robustes et fiables lors de la procédure de cross-validation.*
Note 2 : la plupart du temps, une placette est un outlier *pour plusieurs variables X. Ainsi, un taux de 2 % appliqué aux 9 variables environnementales ne conduit pas à retirer de l'analyse 18 % des placettes : 11-14% selon l'essence.*

4.4 Bilan des placettes disponibles par essence et type de modèle

Distribution

Essence	Nb. placettes	Variables
Abies alba	33445	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof
Fagus sylvativa	33445	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof
Picea abies	33445	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof
Pinus sylvestris	33445	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof
Quercus petraea	33445	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof
Quercus pubescens	33445	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof
Quercus robur	33445	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof

Productivité

Essence	Nb. placettes	Variables[1]
Abies alba	349	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof
Fagus sylvativa	422	Twin, Tsum, Pwin, Psum, VPDsum, pH, RUM et Prof
Picea abies	498	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof
Pinus sylvestris	640	Twin, Tsum, Pwin, Psum, VPDsum, pH, RUM et Prof
Quercus petraea	522	Twin, Tsum, Pwin, Psum, VPDsum, pH, RUM et Prof
Quercus pubescens	136	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof
Quercus robur	336	Twin, Tsum, Pwin, Psum, VPDsum, pH, CN, RUM et Prof

Tableau 7. *Bilan des espèces, nombre d'observation, et variables prédictrices retenues pour la modélisation de la productivité et de la probabilité de présence des espèces forestières.*
[1] On rappelle que suivant les espèces, l'analyse des corrélations pH/C:N a pu conduire à sélectionner le pH comme prédicteur prioritaire.

5 Modélisation

5.1 Modèle additif généralisé

Le cadre de modélisation statistique retenu est le **modèle additif généralisé (GAM)**. Il s'agit d'un cadre non paramétrique qui associe des propriétés du modèle linéaire généralisé avec celles du modèle additif. Le GAM permettent de détecter les effets non-linéaires des prédicteurs sur la variable X et de se limiter aux effets propres de ces prédicteurs (interactions non testées). L'ajustement d'un GAM nécessite le choix d'une fonction de lien et d'un type de lissage.

- **Fonction de lien** : elle dépend de la nature de la variable Y. Dans le cas des modèles de distribution, le modèle d'erreur est **binomial** ; dans le cas des modèles de productivité, le modèle d'erreur est **gaussien**.
- **Lissage** : il dépend de la nature des relations entre Y et [X_1 ; X_2 ; … ; X_n] (grandes tendances = lissage fort ; ou variations locales autour de la tendance = lissage faible). Nous sommes ici intéressés par les grandes tendances, les fluctuations locales pouvant être d'ordre artéfactuel. Il convient ici de choisir une méthode de lissage rigide. De nombreuses techniques ont été comparées par simulation pour obtenir le meilleur compromis biais-variance (lœss vs. spline, différents paramétrages de la spline ou du lœss ; **Mérian 2013**). Il en résulte que les techniques donnent des résultats très similaires, et d'autant plus similaires que le lissage est rigide. Il a été retenu un lissage par **lœss de degré 1 et de fenêtre de largeur (SPAN) 0.7 (c'est-à-dire une largeur égale à 70 % de l'amplitude des données)**.

5.2 Validation croisée et statistiques prédictives

Les modèles servant à faire des projections dans le futur, une procédure de cross-validation a été réalisée lors de l'ajustement estimer la fiabilité et la capacité prédictive. Selon le type de modèle (distribution ou productivité), le nombre de fractions N de cross-validation, le nombre de répétitions P de la procédure de cross-validation et les statistiques varient (cf. tableau en partie 5.6).

Le N-partitionnement du jeu de données conduit à avoir, de temps en temps, des placettes de la fraction de validation en dehors du domaine de calibration (i. e. du domaine défini par les N-1 fractions de calibration). La mauvaise capacité des GAM à extrapoler conduit à obtenir des résultats d'autant plus différents entre deux procédures successives de cross-validation que les placettes de validation sont en dehors du domaine de calibration. Deux solutions pour limiter ce problème : (1) augmenter le nombre de fractions et, (2) effectuer P cross-validations et prendre la valeur **médiane** des estimations des statistiques prédictives (dans ce cas, les fractions de validation-croisées ont été obtenues par tirage aléatoire, sans chevauchement entre fractions). Ces solutions sont couteuses en temps.

- Distribution, le risque d'avoir une placette fortement en dehors du domaine de calibration est quasi-nul, en lien avec le nombre de placettes. De plus, les prédictions étant bornées entre 0 et 1, la mauvaise extrapolation des GAM est également bornée.
-
- Productivité, le risque est élevé (faible nombre de placettes et prédictions variant entre [0 ; +∞[). A la suite de tests, il a été décidé de fixer N = 10 et P = 10. Ces valeurs sont un bon compromis entre temps de calcul et justesse des statistiques prédictives (la valeur vraie de la statistique étant obtenue par une cross-validation en *leave-one-out*).

5.3 Modèle nul

Le modèle nul est le modèle de base auquel seront confrontés les modèles avec des variables environnementales.

- Distribution : le modèle nul est un modèle à une variable X, avec X une variable aléatoire suivant une loi normale ;
- Productivité : le modèle nul est un modèle à deux variables X : (H0 ; RDI). Le modèle à deux variables est d'abord ajusté ; le poids respectif de chacune des variables est ensuite testé en comparant les modèles emboités : (1) 'H0' et 'H0 + RDI' pour estimer le poids de RDI, et (2) 'RDI' et 'H0 + RDI' pour estimer le poids de H0. Selon l'essence, les variables retenues sont soit RDI soit H0 + RDI.

5.4 Intégration de nouvelles variables, comparaison de modèles emboîtés

Pour une étape E de la construction du modèle, l'ajout de chaque variable environnementale non intégrée dans le modèle à l'étape précédente E-1 est testé. Le nouveau modèle est comparé au modèle de l'étape E-1. Une variable environnementale est intégrée SSI :

- amélioration de la capacité prédictive du modèle ;
- cohérence de son effet propre (bibliographie) ;
- conservation des effets propres des variables déjà intégrées.

Dans le cas où plusieurs variables environnementales remplissent ces critères, les conditions suivantes s'appliquent **dans l'ordre** :

- choisir la variable climatique si le modèle n'en comporte pas ;
- choisir la variable qui améliore le plus la capacité prédictive du modèle.

5.5 Importance de chaque variable

Une fois le modèle construit, l'importance de chaque variable est estimée en comparant le modèle complet au modèle dans lequel la variable pour laquelle on souhaite estimer l'importance est retirée.

5.6 Tableau récapitulatif

Modèle	Distribution	Productivité
Modèle d'erreur	Binomial	Gaussien
Modèle nul	Variable aléatoire	H0 + RDI
Statistiques de prédiction	AUC	R^2, variance des erreurs
N (fractions)	3	10
P (répétitions)	1	10
Seuil d'intégration	$\Delta_{AUC} > 0.01$	*p*-value < 0.05 (test de Pitman-Morgan sur la variance des erreurs ; **Morgan 1939**)

Tableau 8. *Caractéristiques des procédures statistiques de sélection de variables dans les modèles de productivité et de probabilité de présence.*

5.7 Packages R

- Ajustement des GAM : le package 'gam' (version 1.09 du package, automne 2013, **Hastie et Tibshirani 1990**) a été choisi. Ce package présente l'avantage d'ajuster simultanément les effets des prédicteurs; l'ajustement GAM obtenu est donc indépendant de l'ordre d'introduction des variables dans le modèle.
- Cross-validation : cette procédure a été programmée par Pierre Mérian ;
- Calcul de l'AUC : package 'pROC' ;
- Test de Pitman-Morgan : package 'PairedData'.

6 Modèles finaux

Les tableaux ci-dessous résument les variables des modèles de distribution et de productivité, ainsi que la forme de l'effet (les graphiques de ces effets sont présentés par essence et par type de modèles – distribution ou productivité – en appendice 2) :

- + : effet monotone positif ;
- +0 : effet positif puis saturant ;
- 0+ : effet plat puis positif ;
- - : effet monotone négatif ;
- -0 : effet négatif puis saturant ;
- 0- : effet plat puis négatif ;
- + - : effet en cloche (admettant un maximum) ;
- - + : effet en cloche inversée (admettant un minimum).

6.1 Distribution

Essence	pH	CN	RUM	Prof	Twin	Tsum	Pwin	Psum	VPDSum	R2-p[1]
A. alba		+ -			+ -			0+		0.441
F. sylvativa	0-					+ -		+0		0.435
P. abies	-					0-		0+		0.512
P. sylvestris	0+	+0			+ -		-			0.394
Q. petraea					+ -		-0		0-	0.449
Q.pubescens				-				+ -	+0	0.476
Q. robur			+	+	+0					0.508

[1] R2 de prédiction, issu de la procédure de validation croisée

6.2 Productivité

Essence	H0	RDI	pH	CN	RUM	Prof	Twin	Tsum	Pwin	Psum	VPDSum	AUC[2]
A. alba	-	+						+		+		
F. sylvativa	-	+	0+			+				+0		
P. abies	-	+					+0					
P. sylvestris		+	-				+ -				-	
Q. petraea	-	+	0-							**0-** **(0.07)**		
Q. pubescens		+						0+		+0		
Q. robur	-	+				**+** **(0.11)**						

[2] AUC (Area Under the R-O Curve), de prédiction, issu de la procédure de validation croisée

Tableau 9. *Effets statistiques et capacité prédictive associés aux modèles ajustés.*

Note : à la vue des difficultés pour intégrer des variables environnementales dans les modèles de productivité, le seuil d'intégration d'une nouvelle variable (p-value du test de Pitman-Morgan < 0.05)

a été relâché. Les variables concernées sont en gras dans le tableau ci-dessus, avec la p-value du test entre parenthèses.

7 Projections

Le tableau suivant présente les périodes de références relatives aux données IFN, aux calibrations, et aux simulations.

	Millésime des données source	Période calibration	Projection « période présente »	Projection future
Principe	5 fractions IFN 2005-2009		Une période climatique unique	71 périodes glissantes médianes 2015 à 2085
Productivité	Périodes de 5 ans, 2000-2004 à 2004-2008	Idem, SAFRAN	2004-2008, SAFRAN	2013-2017 à 2083-2087, CERFACS scratch 2010
Distribution	Constatée sur 2005-2009	1971-2000 (climat trentenaire), SAFRAN	1980-2009, SAFRAN	2001-2030 à 2071-2100, CERFACS scratch 2010

Tableau 10. *Périodes de référence et millésimes relatifs aux données IFN modélisées, aux calibrations, et aux projections (période présente et future).*

Quelles combinaisons « GCM x Scénario x Membre » ?

Les modèles de distribution et de productivité ont été projetés à l'horizon 2100 pour les 15 combinaisons « GCM x Scénario x Membre » fournies par le CERFACS. Les variables édaphiques sont supposées stables dans le temps, et visent essentiellement à accroître le réalisme spatial des modèles. Pour chaque essence, **H0 et RDI sont fixés à la moyenne nationale calculée sur les placettes utilisées pour la construction des modèles**.

Quelles périodes ?

Les périodes de projection futures sont de même intervalle que la période de calibration : 30 ans pour la distribution et 5 ans pour la productivité. Entre 2001 et 2100, 71 périodes peuvent être définies pour la distribution (2001-2030 à 2071-2100) et 96 pour la productivité (2001-2005 à 2096-2100). **71 années médianes sont donc communes aux deux modèles : de 2015 (2001-2030 pour la distribution et 2013-2017 pour la productivité) à 2085 (2071-2100 pour la distribution et 2083-2087 pour la productivité).**

Bornage des valeurs hors domaine de calibration

Pour un pixel 8km, une période et une combinaison « GCM x Scénario x Membre » donnés, les prédicteurs environnementaux peuvent se retrouver en dehors du domaine de calibration (DC) des modèles. Les pixels avec des données édaphiques hors DC sont identiques dans le temps, alors que

les évolutions climatiques tendent à faire augmenter le nombre de pixels hors DC pour tout ou partie des prédicteurs (hausse des températures et des VPD, et baisse des précipitations).

Les extrapolations des GAM étant hasardeuses, les prédicteurs environnementaux sont bornés au domaine de calibration. Pour chaque essence, type de modèle (distribution ou productivité) et variable, le caractère « hors DC / DC » de chaque pixel est enregistré.

8 Calcul des modificateurs

8.1 Principe

Il s'agit ici de modificateurs des taux de passage entre classes de diamètre du modèle LSFDM, qui est calibré par région administrative et par groupe d'essences (feuillus / résineux). Ces modificateurs (**Bontemps 2012**) quantifient la variation relative dans le temps de la productivité des forêts par rapport à la productivité actuelle. La productivité des forêts doit donc être estimée à t_0 et à t, le modificateur étant **le rapport de la productivité à t par groupe d'essences sur la productivité par groupe d'essences à t_0.**

Par région, la productivité d'un groupe d'essences (feuillus/résineux) à t correspond à la moyenne des productivités de chaque essence pondérée par les fréquences d'essence. **La fréquence à t d'une essence est elle-même définie comme la fréquence à t_0 (définie à partir des fractions récentes de l'IFN), multipliée par le rapport entre les probabilités de présence à t et t_0.** On utilise donc la variation relative des probabilités de présence.

Pour calculer les modificateurs, il faut donc calculer (1) la fréquence des essences par région à t_0, (2) la probabilité de présence des essences par région à t_0, (3) la probabilité de présence des essences par région à t, (4) la productivité des essences par région à t.

8.2 Calcul des modificateurs

8.2.1 Nomenclature

Nous conservons des terminologies identiques à celles utilisées lors des présentations de référence données dans l'ANR Oracle (**Mérian et Bontemps, 2013**).

Soit e une essence parmi une liste de E éléments (E = 7).
Soit r une région parmi une liste de R éléments (R = 22).
Soit t une période définie par son année médiane parmi une liste de T éléments (T = 71 ; 2015 à 2085).
Soit f une fréquence. On note $f_{e,r,t}$ la fréquence de l'essence e dans la région r à la période t.
Soit P une probabilité de présence. On note $P_{e,r,t}$ la probabilité de présence de l'essence e dans la région r à la période t.

Soit G une productivité. On note $G_{e,r,t}$ la productivité de l'essence e dans la région r à la période t.

8.2.2 État initial de la forêt

Calcul des fréquences d'essences régionales

Les fréquences initiales sont calculées sur le champ IFN « essence principale ». Le calcul est donc réduit aux placettes IFN pour lesquelles ce champ est renseigné (28206 placettes).

Soit $N_{e,r,0}$ le nombre de placettes IFN dans la région r où l'essence principale est l'essence e. Cette donnée est **observée** à la période initiale (t = 0). La fréquence d'une essence e dans une région r est définie comme :

$$f_{e,r,0} = \frac{N_{e,r,0}}{\sum_{e=1}^{E} N_{e,r,0}} \quad \textbf{(1)}$$

Calcul des fréquences d'essences régionales normées

La somme des fréquences initiales des 7 essences objectives est inférieure à 1 car d'autres essences sont présentes dans les régions. La forêt modélisée se limitant aux 7 essences objectives, la somme des fréquences initiales doit valoir 1 par région. **On décide donc de « normer » ces fréquences**. Si une essence parmi les 7 analysées est absente d'une région ($f_{e,r,0} = 0$), sa fréquence est remplacée par un « germe » $g_{e,r,0}$ dont la valeur est définie pour que la **fréquence normée valle 0.001**. **Ce germe permet d'initier un processus de colonisation éventuel qui proviendrait de l'évolution favorable de la probabilité de présence de l'espèce actuellement absente dans cette région**. Pour une région r donnée**, le germe** $g_{e,r,0} = f_{e,r,0}$ **est calculé comme suit** :

$$g_{e,r,0} = \frac{\sum_{e=1}^{E} \left(f_{e,r,0} <> 0 \right)}{1 - p_r \times 0.001} \quad \textbf{(2)}$$

avec p_r < 7, le nombre d'essences absentes de la région r. Il faut en effet que $g_{e,r,0}$ / $\sum_{e=1}^{E} \left(f_{e,r,0} \right)$ = $g_{e,r,0}$ / ($\sum_{e=1}^{E} \left(f_{e,r,0} <> 0 \right)$+ p_r $g_{e,r,0}$) = 0.001, d'où vient le résultat.

Les fréquences initiales normées F sont donc calculées comme suit :

$$F_{e,r,0} = \frac{f_{e,r,0}}{\sum_{e=1}^{E} f_{e,r,0}} \quad \textbf{(3)}$$

On dispose donc d'un tableau des $F_{e,r,0}$ de 22 lignes (régions) et 7 colonnes (essences). **Fichier TXT correspondant : Freq0_sp.txt.**

Estimation des probabilités initiales

Pour estimer la probabilité de présence par essence et par région à t = 0, les modèles de **distribution sont projetés sur la France avec les données climatiques de la période 1980-2009, période la plus récente possible avec les données SAFRAN**. On obtient donc une probabilité initiale **modélisée** pour chaque essence et pixel de 8 km. La moyenne $P_{e,r,0}$ et l'écart-type (demande du LEF pour ce dernier) sont ensuite calculées pour chaque essence à l'échelle de la région. On dispose donc du tableau des $P_{e,r,0}$ et du tableau des écart-types, chacun de 22 lignes (régions) et 7 colonnes (essences). **Fichiers TXT correspondants : Proba0_sp_M.txt et Proba0_sp_SD.txt.**

Estimation des productivités initiales

La démarche est similaire à celle de l'estimation des probabilités initiales, **mais elle fait intervenir une pondération par la probabilité de présence** : les modèles de productivité sont projetés sur la France au pixel de 8km avec **les données climatiques de la période 2004-2008**. On obtient donc une productivité initiale **modélisée**. Ces productivités sont ensuite agrégées pour chaque essence à l'échelle de la région en calculant une moyenne pondérée par la probabilité de présence en tout point, $G_{e,r,0}$. On dispose donc du tableau des $G_{e,r,0}$ et du tableau des écart-types, chacun de 22 lignes (régions) et 7 colonnes (essences). **Fichiers TXT correspondants : PRODP0_sp_M.txt et PRODP0_sp_SD.txt.**

A noter qu'à des fins comparatives, une version non-pondérée de la productivité spécifique régionale a également été calculée. Fichiers TXT correspondants : Prod0_sp_M.txt et Prod0_sp_SD.txt.

Estimation de la productivité régionale initiale des groupes d'essences

Deux groupes d'essences sont définis (feuillus/résineux). Le calcul suivant vaut pour un groupe C composé de E' essences. La productivité de ce groupe dans une région r donnée correspond à la moyenne des productivités de chaque essence, pondérée par les fréquences de ces mêmes essences. Elle est définie comme :

$$G_{C,r,0} = \frac{\sum_{e=1}^{E'} \left(F_{e,r,0} \times G_{e,r,0} \right)}{\sum_{e=1}^{E'} F_{e,r,0}} \qquad \textbf{(4)}$$

On dispose donc d'un tableau des $G_{C,r,0}$ de 22 lignes (régions) et 2 colonnes (groupes). **Fichier TXT correspondant : PRODP0_gr.txt.**

A noter qu'à des fins comparatives, une version utilisant les productivités spécifiques régionales non pondérées a également été calculée. Fichiers TXT correspondants : Prod0_gr.txt.

8.2.3 Estimation de l'état futur des forêts

Estimation des probabilités à la période t

La démarche est identique à celle de l'estimation des probabilités initiales : les modèles de distribution sont projetés sur la France au pixel de 8 km avec les données climatiques de la période t. On obtient donc une probabilité **modélisée** pour chaque combinaison « Période t x GCM x Scénario x Membre ». Ces données sont ensuite agrégées pour chaque essence à l'échelle de la région en calculant la probabilité moyenne $P_{e,r,t}$ et son écart-type. Pour chaque combinaison « Période t x GCM x Scénario x Membre », on dispose donc du tableau des $P_{e,r,t}$ et du tableau des écart-types, chacun de 22 lignes (régions) et 7 colonnes (essences). **Fichiers TXT correspondants : Proba_sp_M.txt et Proba_sp_SD.txt.** *Note : ces fichiers sont la concaténation des données sur l'ensemble des périodes x GCM x Scénario x Membres.*

Estimation des fréquences à la période t

La fréquence $F_{e,r,t}$ des essences à la période t est une modification de la fréquence initiale par le ratio des probabilités de présence entre les périodes t et 0. Elle se définie donc comme :

$$F_{e,r,t} = F_{e,r,0} \times \frac{P_{e,r,t}}{P_{e,r,0}} \qquad \textbf{(5)}$$

Lorsque la somme des fréquences des essences d'une région r donnée est supérieure à 1 ($\sum_{e=1}^{E} F_{e,r,t} > 1$), les fréquences sont normées pour que la somme valle 1. **Notons qu'une baisse de la somme des fréquences traduit un recul de la forêt dans la région (forêt réduite aux E essences), et dans ce dernier cas, on ne norme pas ces fréquences, qui traduisent alors un déclin de la forêt (c'est donc un indicateur implicite de la mortalité, et par conséquent un support pour rendre le LSFDM mortalité-dépendant, au plan de la causalité climatique, comme on le précise plus loin).**

Pour chaque combinaison « Période t x GCM x Scénario x Membre », on dispose donc d'un tableau des $F_{e,r,t}$ de 22 lignes (régions) et 7 colonnes (essences). **Fichier TXT correspondant : Freq_sp.txt.** *Note : ce fichier est la concaténation des données sur l'ensemble des périodes x GCM x Scénario x Membres.*

Estimation des productivités à la période t

La démarche est identique à celle de l'estimation des productivités initiales : les modèles de productivité sont projetés sur la France au pixel de 8 km avec les données climatiques de la période t. On obtient donc une productivité **modélisée** pour chaque combinaison « Période t x GCM x Scénario x Membre ». Ces données sont ensuite agrégées pour chaque essence à l'échelle de la région en calculant une productivité moyenne pondérée par la probabilité de présence en tout point $G_{e,r,t}$ et son écart-type. Pour chaque combinaison « Période t x GCM x Scénario x Membre », on dispose donc du tableau des $G_{e,r,t}$ et du tableau des écart-types, chacun de 22 lignes (régions) et 7 colonnes (essences). **Fichiers TXT correspondants : PRODP_sp_M.txt et PRODP_sp_SD.txt.** *Note : ces fichiers sont la concaténation des données sur l'ensemble des périodes x GCM x Scénario x Membres.*

A noter qu'à des fins comparatives, une version non-pondérée de la productivité spécifique régionale a également été calculée. Fichiers TXT correspondants : Prod_sp_M.txt et Prod_sp_SD.txt.

Estimation de la productivité régionale des groupes d'essences à période t

La démarche est identique à celle de l'estimation de la productivité régionale initiale des groupes d'essences. Dans une région r et pour un groupe C composé de E' essences, la productivité à la période t est définie comme :

$$G_{C,r,t} = \frac{\sum_{e=1}^{E'} \left(F_{e,r,t} \times G_{e,r,t} \right)}{\sum_{e=1}^{E'} F_{e,r,t}} \qquad \textbf{(6)}$$

A noter ici que cette productivité apparaît comme étant corrigée de l'éventuelle baisse de la somme des fréquences des espèces dans le temps (situation de « déclin » des espèces, cf supra). Sa variation dans le temps ne mesure donc que l'effet du climat se traduisant sur la productivité des espèces, pas sur leur fréquence.

Pour chaque combinaison « Période t x GCM x Scénario x Membre », on dispose donc d'un tableau des $G_{C,r,t}$ de 22 lignes (régions) et 2 colonnes (groupes). **Fichier TXT correspondant : PRODP_gr.txt.** *Note : ce fichier est la concaténation des données sur l'ensemble des périodes x GCM x Scénario x Membres.*

A noter qu'à des fins comparatives, une version utilisant les productivités spécifiques régionales non pondérées a également été calculée. Fichiers TXT correspondants : Prod_gr.txt.

Estimation des modificateurs à période t

Pour un groupe C et une région r, le modificateur M des taux de passage du modèle de dynamique de la ressource est défini comme **le ratio des productivités entre les périodes t et 0 multiplié par le ratio des fréquences entre les périodes t et 0** :

$$M_{C,r,t} = \frac{G_{C,r,t}}{G_{C,r,0}} \times \frac{\sum_{e=1}^{E'} F_{e,r,t}}{\sum_{e=1}^{E'} F_{e,r,0}} \qquad \textbf{(7)}$$

Pour comprendre ce choix de formulation, il faut remarquer que le modificateur M s'écrit aussi sous la forme suivante :

$$M_{C,r,t} = \frac{\dfrac{\sum_{e=1}^{E'} \left(F_{e,r,t} \times G_{e,r,t}\right)}{\sum_{e=1}^{E'} F_{e,r,t}} \times \sum_{e=1}^{E'} F_{e,r,t}}{\dfrac{\sum_{e=1}^{E'} \left(F_{e,r,t0} \times G_{e,r,t0}\right)}{\sum_{e=1}^{E'} F_{e,r,t0}} \times \sum_{e=1}^{E'} F_{e,r,t0}} = \frac{\sum_{e=1}^{E'} \left(F_{e,r,t} \times G_{e,r,t}\right)}{\sum_{e=1}^{E'} \left(F_{e,r,0} \times G_{e,r,0}\right)} = \frac{G'_{C,r,t}}{G'_{C,r,t0}} \qquad \textbf{(8)}$$

Il est donc identique à celui qui serait défini par un ratio de productivités qui tiendraient compte, à la fois des variations de productivité spécifiques, et des variations de fréquences spécifiques (G'). A ce titre, le modificateur ainsi défini, appliqué aux taux de croissance diamétriques du modèle de ressource, est bien un moyen de faire varier la taille de la population d'arbres au cours du temps en fonction de la productivité, mais aussi d'une grandeur qu'on peut interpréter un phénomène de mortalité. La formulation (7) présente alors l'avantage d'offrir une décomposition du modificateur, selon ses composantes liées à la fréquence des espèces, et à leur productivité.

Pour chaque combinaison « Période t x GCM x Scénario x Membre », on dispose donc d'un tableau des $M_{C,r,t}$ de 22 lignes (régions) et 2 colonnes (groupes). **Fichier TXT correspondant : MODP_gr.txt.** *Note : ce fichier est la concaténation des données sur l'ensemble des périodes x GCM x Scénario x Membres.*

A noter qu'à des fins comparatives, une version utilisant les productivités spécifiques régionales non pondérées a également été calculée. Fichiers TXT correspondants : Modif_gr.txt.

9 Livrables

Les livrables sont fournis sous plusieurs format : tables, rasters, Rdata et graphiques.

9.1 Tables et Rdata

L'ensemble des tables sont fournies sous format TXT, séparateur tabulation. Elles fournissent les variables **par région**.

Tables relatives à l'état initial

- FREQ0_sp : fréquences initiales par essence et par région ;
- PROBA0_sp_M : probabilités initiales moyennes par essence et par région;
- PROBA0_sp_SD : écart-types régionaux des probabilités initiales par essence;

- PROD0_sp_M : productivités initiales moyennes par essence et par région;
- PROD0_sp_SD : écart-types régionaux initiaux des productivités par essence;
- PROD0_gr : productivités initiales moyennes par groupe d'essences et région;

- PRODP0_sp_M : productivités initiales en moyenne pondérée (par la probabilité de présence) par essence et par région;
- PRODP0_sp_SD : écart-types régionaux pondérés (par la probabilité de présence) des productivités initiales par essence et région.
- PRODP0_gr : productivités initiales en moyenne pondérée (par la probabilité de présence) par groupe d'essences et région ;

Tables relatives aux prédictions (une table est la concaténation des données sur l'ensemble des périodes x GCM x Scénario x Membres)

- PROBA_sp_M : probabilité de présence moyenne des essences, par région, période, GCM et scénario climatique ;
- PROBA_sp_SD : écart-type de la probabilité de présence par région (0 à 1), essence, période, GCM et scénario climatique ;
- PROBA_sp_OUT : proportion régionale des pixels hors domaine de calibration des modèles de probabilités, par essence, région, variable, période, GCM et scénario climatique. 0 : aucune cellule en dehors du domaine de calibration ; -1 : variable non présente dans le modèle de l'essence considérée ;
- FREQ_sp : fréquences relatives des essences par région ;

- PROD_sp_M : productivités moyennes par essence et par région;
- PROD_sp_SD : écart-types régionaux des productivités par essence;

- PROD_sp_OUT : proportion régionale des pixels hors domaine de calibration des modèles de productivité, par essence, région, variable, période, GCM et scénario climatique (même codage) ;
- PROD_gr : productivités moyennes par groupe d'essences (feuillus/résineux) et régions ;
- MOD_gr : modificateurs régionaux et par groupe d'essences (feuillus/résineux) des taux de passage du modèle de dynamique forestière (variation relative de la productivité moyenne par groupe d'essences) ;

- PRODP_sp_M : productivités en moyenne pondérée (par la probabilité de présence) par essence et par région;
- PRODP_sp_SD : écart-types régionaux pondérés (par la probabilité de présence) des productivités par essence et région.
- PRODP_gr : productivités en moyenne pondérée (par la probabilité de présence) par groupe d'essences (feullus/résineux) et région ;
- MODP_gr : modificateurs régionaux et par groupe d'essence (feuillus/résineux) des taux de passage du modèle de dynamique forestière (variation relative de la productivité en moyenne pondérée par groupe d'essences) ;

Rdata

Le fichier ORACLE.RData compile l'ensemble des tables présentées ci-dessus. Les données sont organisées sous forme de listes de tables pour faciliter leur utilisation. Les données sont regroupées en deux objets :

- Tables.actuel : ensemble des données relatives à l'état initial ;
- Tables.pred : ensemble des données relatives aux prédictions.

9.2 Rasters

Les rasters fournissent les prédictions des modèles de distribution et de productivité au pixel de 8km pour les périodes [2015 ; 2025 ; … 2075 ; 2085] (8 périodes au total). Un raster est ainsi fourni au format ***.gri*** pour chaque combinaison « Type de modèle x essence x GCM x Scénario x Membre x Période », soit un total de 1680 rasters.

Le nom de chaque fichier renseigne sur son contenu puisqu'il est construit comme suit :

type de modèle_essence_GCM_scénario ou membre_période.gri

Le code d'extraction est fourni à l'**appendice 3**.

Dans un fichier donné, la première colonne contient les valeurs prédites. Les autres colonnes codent pour chacune des variables du modèle : 0 lorsque le pixel est hors domaine de calibration, 1 sinon (la raison est qu' les calculs sur les couches de type somme, moyenne, etc. sont plus faciles avec ce codage binaire).

9.3 Graphiques

A chaque raster présenté ci-dessus est associée une carte de la France sur laquelle sont projetées les données. Les cellules sur-grisées correspondent aux cellules pour lesquelles au moins une variable du modèle présente une donnée hors domaine de calibration.

10 Références bibliographiques

AgroParisTech-ENGREF (UMR LERFoB), IFN (2008). Guide d'utilisation de la carte des pH de surface des sols forestiers français. Version 1, mars 2008, selon le document © accord AgroParisTech (UMR LERFoB) – IFN n° 2007-CPA-2-072 dans sa version d'avril 2007.

Bontemps, J.-D. 2012. Oracle work program – team Forest Ecology, LERFoB. 13 juin 2012, rapport technique interne, 14 p.

Bontemps, J.-D. et P. Mérian. 2013. Modéliser la dépendance climatique du modèle de dynamique de la forêt française (LSFDM - FFSM) : hypothèses de modélisation et de simulation, livrables, et points à débattre. Meeting Oracle, Nancy, 26 novembre 2013, 19 diapositives.

Caurla, S., F. Lecocq, P. Delacote et A. Barkaoui. 2010. French Forest Sector Model: version 1.0. Presentation and theorical foundations. Document de travail du LEF n° 2010-04, 1-26.

Charru, M. 2012. La productivité forestière dans un environnement changeant - Caractérisation multi-échelle de ses variations récentes à partir des données de l'Inventaire Forestier National (IFN) et interprétation environnementale, Nancy (France), p.417.

Gégout, J. C., J. C. Hervé, F. Houllier, et J. C. Pierrat. 2003. Prediction of forest soil nutrient status using vegetation. Journal of Vegetation Science 14:55-62.

Hastie, T. et R. Tibshirani. 1990. Generalized additive models. Chapman & al., 352 p.

Le Moigne, P. 2002. Description de l'analyse des champs de surface sur la France par le système SAFRAN. Rapport technique, Centre national de recherches météorologiques, Météo-France, 2002.

Mérian, P. 2013. Effet de la technique de lissage sur le lissage d'un nuage de points – Approche par simulation. Rapport technique interne, 11 p.

Mérian, P. et J.-D. Bontemps. 2013. Modélisation des relations environnement-productivité et environnement-distribution des principales essences françaises. Meeting Oracle, Nancy, 26 novembre 2013, 33 diapositives.

Morgan, W. A. 1939. A test for the significance of the difference between two variances in a sample from a normal bivariate distribution. Biometrika 31:13-19.

Pagé, C. 2008. Format des données SAFRAN et scénarios climatiques désagrégés au CERFACS. Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, Toulouse. p. 10.

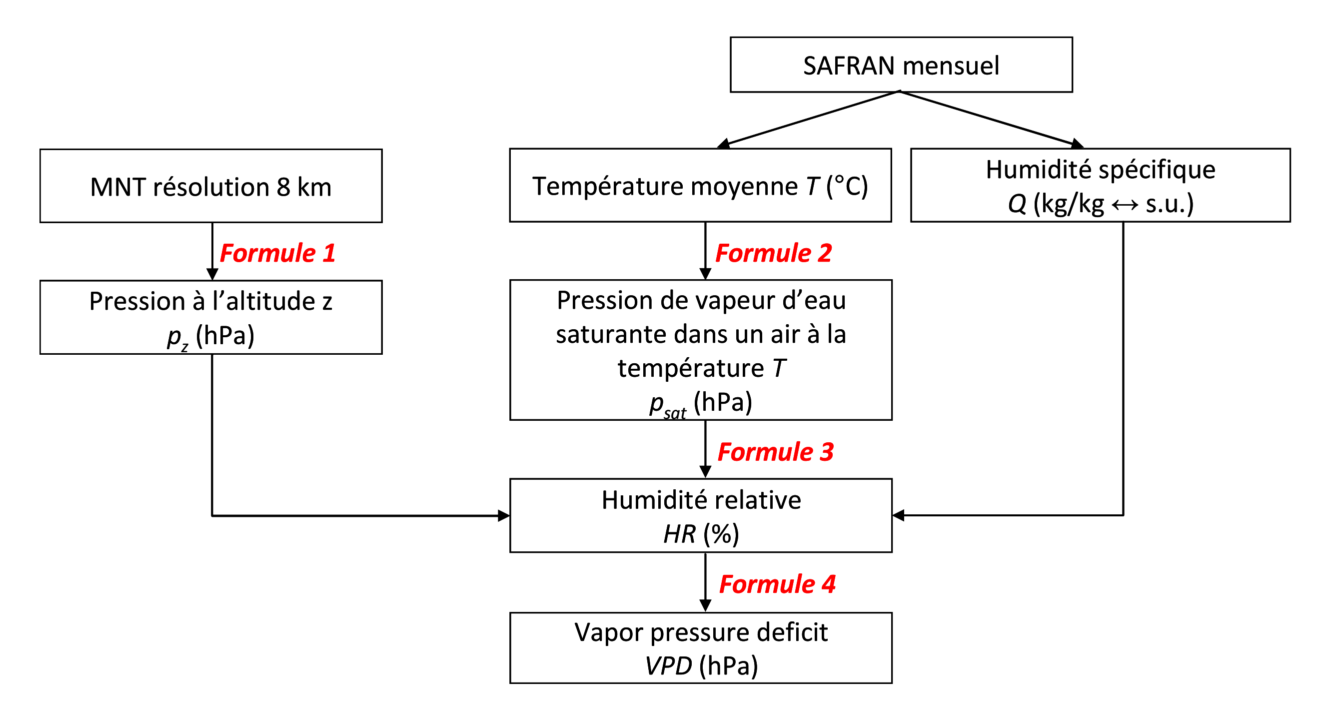
Pagé, C. et L. Terray. 2011. Nouvelles projections climatiques à échelle fine sur la France pour le 21ème siècle : les scénarii SCRATCH2010. Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, Toulouse, p. 25.

Piedallu, C., J.-C. gégout, V. Perez, F. Lebourgeois. 2013. Soil water balance performs better than climatic water variables in tree species distribution modelling. Global Ecology and Biogeography 22, 470-482.

Wernsdorfer, H., A. Colin, J. D. Bontemps, H. Chevalier, G. Pignard, S. Caurla, J. M. Leban, J. C. Hervé, et M. Fournier. 2012. Large-scale dynamics of a heterogeneous forest resource are driven jointly by geographically varying growth conditions, tree species composition and stand structure. Annals of Forest Science 69:829-844.

11 Appendice 1 : calcul du VPD

1. Diagramme de calcul



2. Formules de calcul

Formule 1 : nivellement barométrique

Cette formule permet d'ajuster la pression en fonction de l'altitude.

$$p_z = p_0 \times \left(1 - \frac{0.0065 \times z}{288.15}\right)^{5.255}$$, avec

p_0 la pression en hPa au niveau de la mer, soit 1013.25 ; z l'altitude

Cette formule fait les hypothèses suivantes : (1) la baisse de la température est de 0.65 °C par tranche de 100 m d'latitude, (2) la température au niveau de la mer est de 15 °C (288.15 = 273.15 + 15).

Formule 2 : calcul de la pression de vapeur d'eau saturante p_{sat}

Les équations sont issues de Buck (1981), Eqs. [8]. Notez que l'équation e'_w s'applique quand T est supérieure à 0.01 et e'_i s'applique quand T est inférieure ou égale à 0.01. L'indice *w* vaut pour water et *i* pour ice.

$$e_w = [1.0007 + (3.46 \times 10^{-6} P)] \times 6.1121 \exp\left[\frac{17.502T}{240.97 + T}\right],$$

$$e_i = [1.0003 + (4.18 \times 10^{-6} P)] \times 6.1115 \exp\left[\frac{22.452T}{272.55 + T}\right]. \quad (8).$$

Avec T = température moyenne de l'air en °C et P la pression atmosphérique en hPa (équivalent donc p_z à dans la formule 1).

Formule 3 : formule de conversion de l'humidité spécifique à l'humidité relative

La formule est issue de la publication de Nadeau et Puiggali (1995) :

$HR = \frac{p_z \times Q}{p_{sat} \times (0.622 + Q)}$, avec Q = humidité spécifique et HR variant entre 0 et 1

Formule 4 : formule de calcul du VPD à partir de l'humidité relative.

$VPD = p_{sat} \times (1 - HR)$, HR entre 0 et 1. VPD exprimé en hPa.

3. Références

Buck, A. L. 1981. New Equations for Computing Vapor Pressure and Enhancement Factor. Journal of Applied Meteorology 20:1527-1532.

Nadeau, J.P. et Puiggali J.R. 1995. Séchage: des processus physiques aux procédés industriels, Lavoisier Tech. et Doc., Paris. (ISBN 2-7430-0018-X)

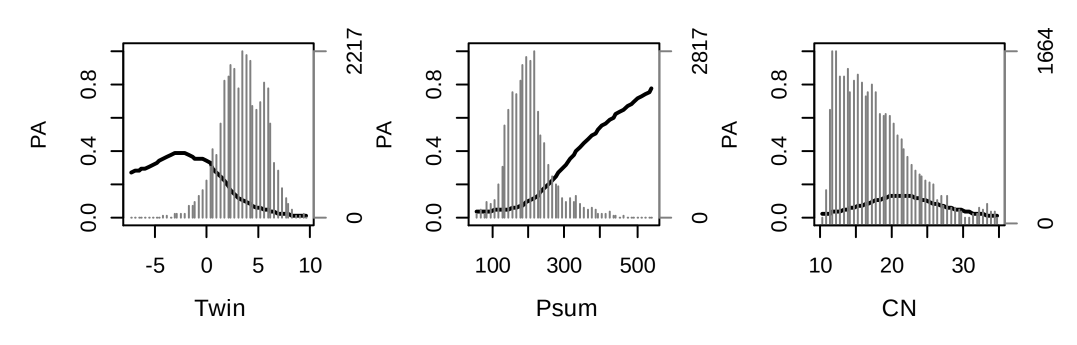
12Appendice 2 : détails des modèles de distribution et de productivité

Les effets des variables sont présentés par essence et type de modèle (distribution ou productivité). Sur chaque graphique, la courbe noire indique l'effet et les bâtons gris la répartition des observations sur le gradient environnemental.

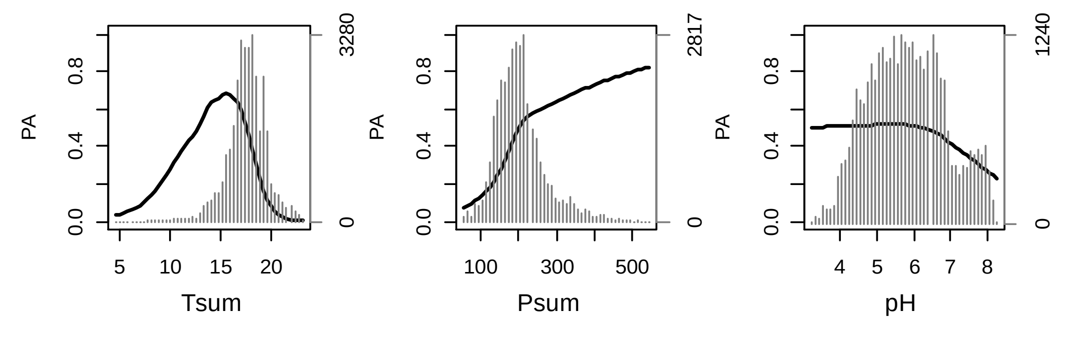
1. Modèles de distribution

La courbe en trait plein noir correspond à la réponse lissée issue de modèles « GAM » (loess gaussien de degré 1, Mérian 2013). La distribution des observations (1/0 indifférenciés) est indiquée par un histogramme.

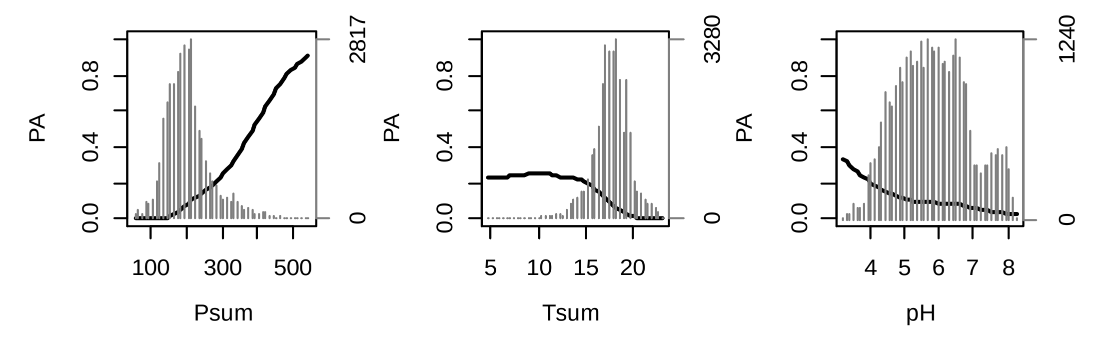
Abies alba



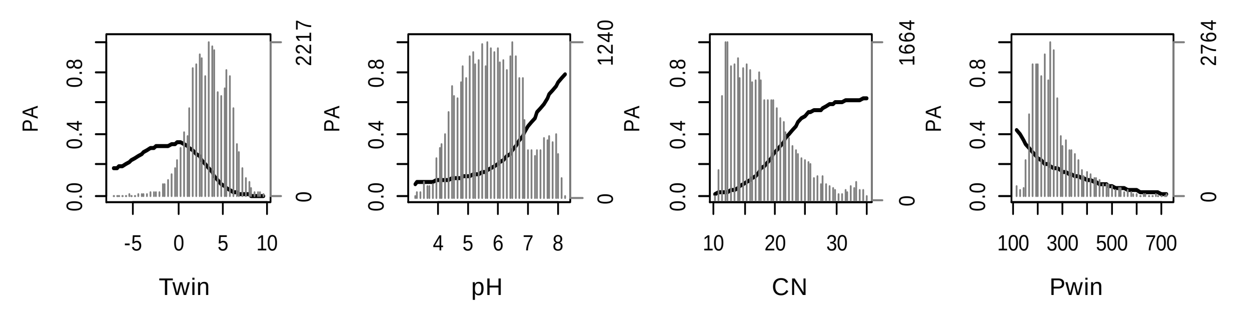

Fagus sylvatica



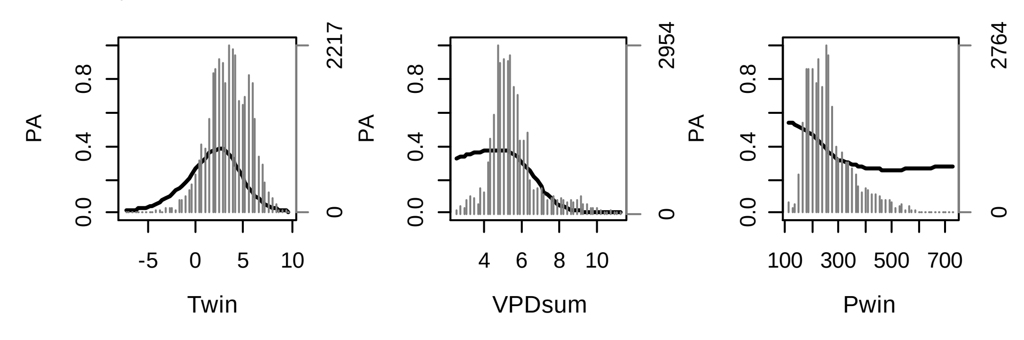
Picea abies



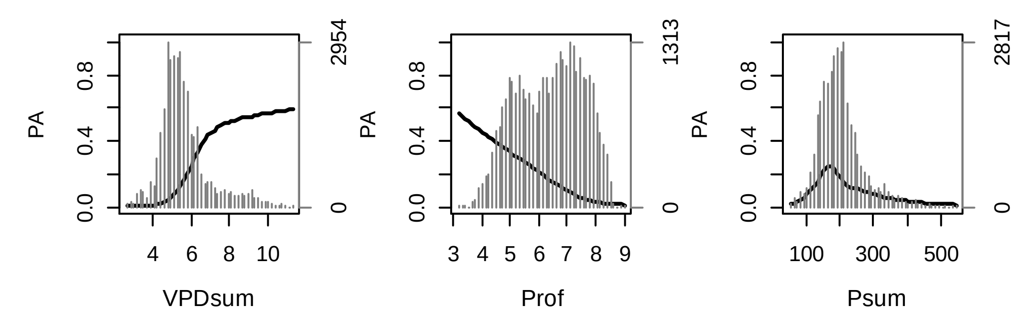
Pinus sylvestris



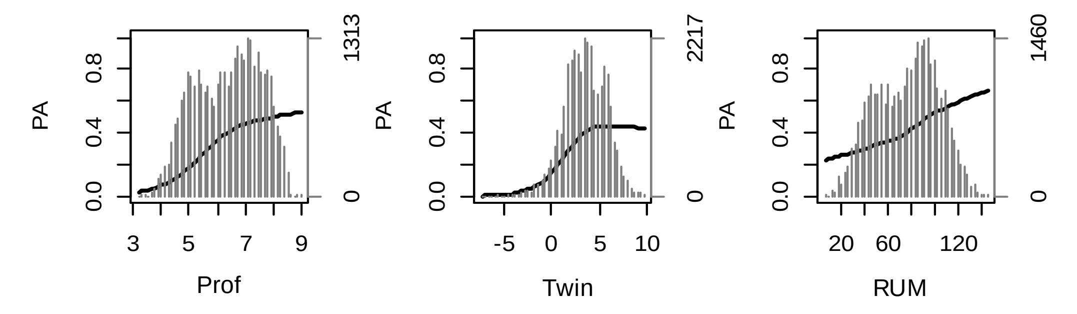
Quercus petraea



Quercus pubescens

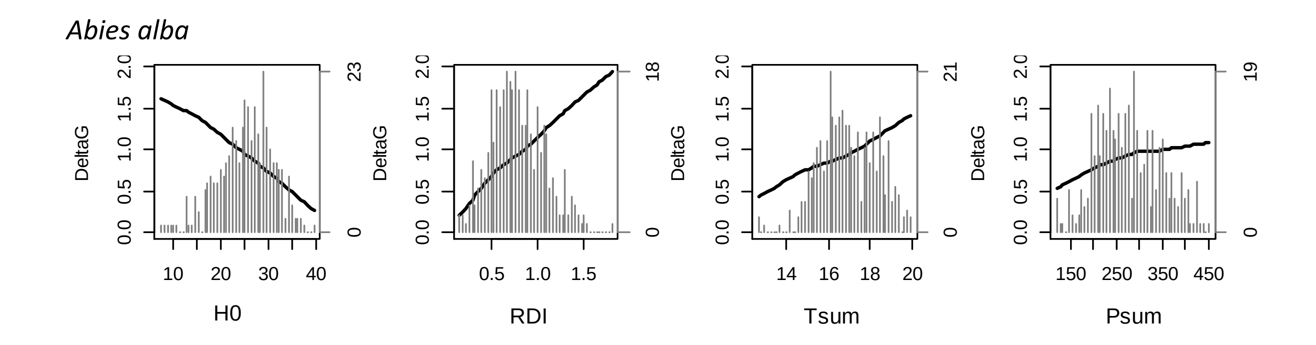


Quercus robur

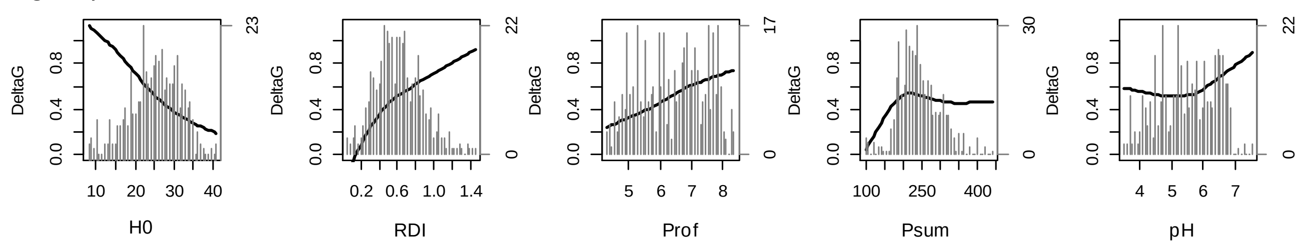


2. Modèles de productivité

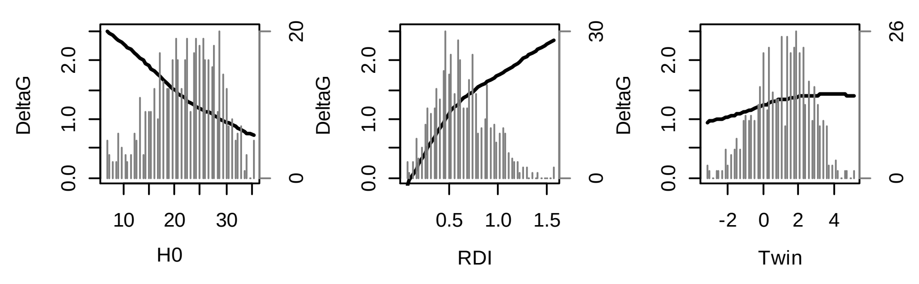
Abies alba



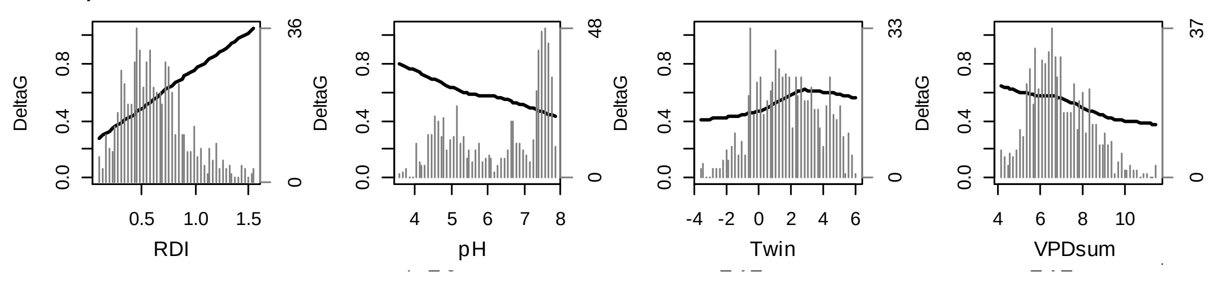
Fagus sylvatica



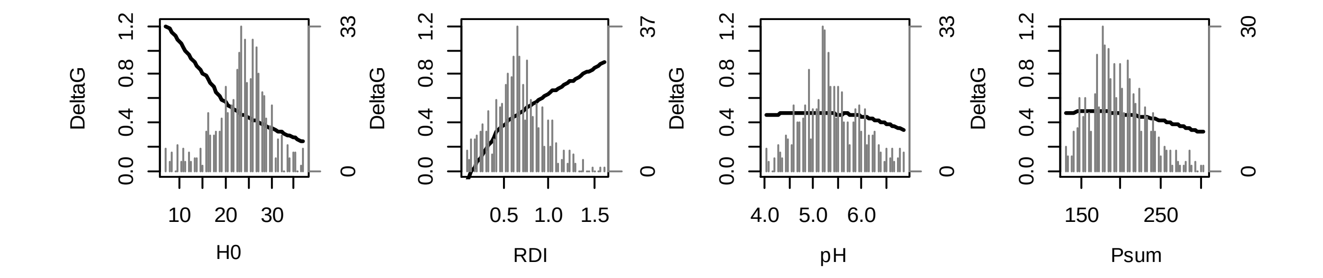
Picea abies



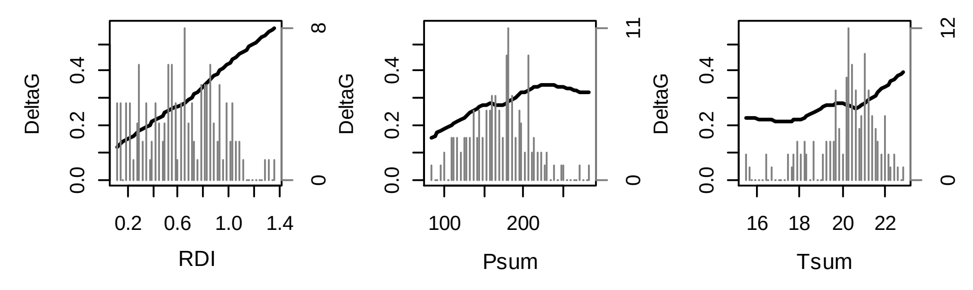
Pinus sylvestris



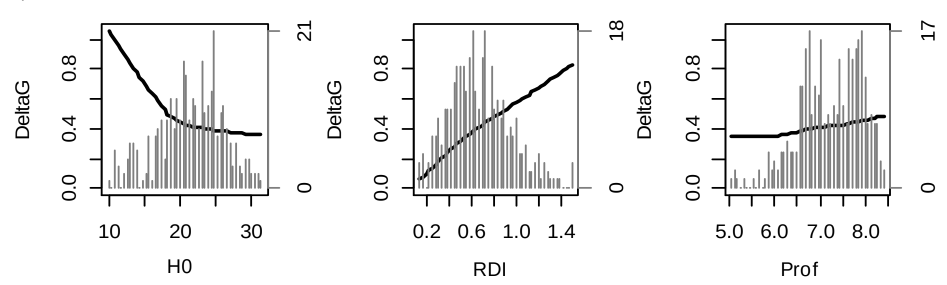
Quercus petraea



Quercus pubescens



Quercus robur



13 Appendice 3

```
library(raster)
REP = "F://Research REF//5 - Post-doc LERFoB//5 - Livrables//Rasters//Productivité//"
NOM = "Prod_Aa_arpege_a1b_2015.grd"
r = raster(paste0(REP, NOM))
```